# Problems and methods in longitudinal research: Stability and change

*Edited by*
DAVID MAGNUSSON,
LARS R. BERGMAN,
*Department of Psychology, University of Stockholm*
GEORG RUDINGER
*Psychologisches Institut, University of Bonn*
*and*
BERTIL TÖRESTAD
*Department of Psychology, University of Stockholm*

EUROPEAN SCIENCE FOUNDATION

*The right of the University of Cambridge to print and sell all manner of books was granted by Henry VIII in 1534. The University has printed and published continuously since 1584.*

CAMBRIDGE UNIVERSITY PRESS
*Cambridge*
*New York Port Chester Melbourne Sydney*

# 9 Application of correspondence analysis to a longitudinal study of cognitive development

JACQUES LAUTREY AND PHILIPPE CIBOIS

Our aim in this chapter is to illustrate in what way the constraints inherent in a specific problem motivate the choice of a particular method to analyze longitudinal research data. The problem examined here is the form of intraindividual variability in level of cognitive development, and stability or changes in this form over time. The method is correspondence analysis. The data used for the purposes of illustration chapter are drawn from a longitudinal study by J. Lautrey, A. de Ribaupierre and L. Rieben. This chapter focuses mainly on methodological issues; more detailed information on the study itself can be found elsewhere, for example in Lautrey, de Ribaupierre & Rieben (1985, 1986), de Ribaupierre, Rieben & Lautrey (1985), Rieben, de Ribaupierre & Lautrey (1983). The first main section of the chapter examines methodological constraints related to the theoretical issues and the nature of the data. The second section is devoted to correspondence analysis and presents the features which make it particularly suited to handling these constraints. The third section deals with the results obtained by correspondence analysis and discusses the implications of some methodological choices.

## CONSTRAINTS INHERENT TO THE NATURE OF THE PROBLEM

### Theoretical issues

The central issue in this study is the form of cognitive development. In other words, does knowledge acquisition adhere to an invariant sequence which is identical for all children, or can cognitive development follow different pathways for different children? In terms of data analysis, such different developmental pathways are inferred from interindividual differences in the form of intraindividual variability. The hypothesis of different pathways refers to the fact that the order of acquisition of two notions, say A and B, can be AB for certain subjects and BA for others. This issue will be examined here by reference to

Piagetian theory, since the postulate of unicity of development is probably formulated most clearly in this view on cognitive development.

According to Piaget, knowledge develops through the construction of mental structures (sensorimotor, concrete, formal) which appear in an invariant sequence. Each of these structures is thought to be general in scope, a feature which is reflected by isomorphism of reasoning across different notional domains at a given point in development. The scope of these structures and their invariant order of construction define a single developmental pathway where the only possible differences between individuals are differences in rate.

The validity of this model has been seriously weakened by the fact that children's level of cognitive development varies widely as a function of the situation in which development is assessed. Showing conclusively that this intraindividual variability corresponds to different pathways in the course of cognitive development rather than to random variations or measurement errors calls for evidence: (1) that intraindividual *décalages* in the order of mastery of various notions do not have the same form (e.g., AB or BA) for different subjects, (2) that they can be accounted for by a meaningful structure at the cross-sectional level, and (3) that such a structure remains stable over time. The third point can only be established through a longitudinal study.

Before presenting a description of this study, the nature of the data calls for discussion.

## Nature of the data

*Subjects.* 154 children were evaluated twice at a three-year interval. They were between the ages of 6 and 12 on the first evaluation (the sample was composed of 22 subjects per age group) and thus between the ages of 9 and 15 on the second evaluation. Since the tasks described below only discriminate ages 6 to 12, only subjects who were between 9 and 12 at the time of the second evaluation (i.e., between the ages of 6 and 9 when tested first) were included in the longitudinal sample. Note, however, that the entire sample was used for the cross-sectional study on the data obtained for the first evaluation. Of the 88 subjects aged 6–9 on the first evaluation, 76 were relocated three years later, thus yielding a 14% loss of subjects.

*Variables.* Subjects were individually administered eight operational tasks adapted from Piaget and Inhelder. Testing adhered as closely as possible to the 'critical questioning' technique developed by Piaget and colleagues. Limited space prevents us from providing a detailed description of these tasks. A brief description of the material, instructions and scoring criteria can be found in Rieben *et al.* (1983). A more succinct version is included in Lautrey *et al.* (1985) or in de Ribaupierre *et al.* (1985). For the present purposes the names of the tasks are provided and indications as to which of the four broad fields of knowledge they are associated with:

Logicomathematical domain:
- class intersection (6 items)
- quantification of probabilities (7 items)

Physics domain:
- conservation (4 items)
- islands (3 items)

Spatial domain:
- sectioning of volumes (5 items)
- unfolding of volumes (3 items)

Mental imagery:
- folding of lines (4 items)
- folds and holes (6 items)

Each of the eight tasks reflects a cognitive operation, and mastery on the task is considered to be indicative of concrete operations. In addition, however, several items in each task tap a given operation in a variety of situations and are thus measures of potential *décalages* in its construction. In total subjects were tested on 38 items on two occasions. For reasons which are discussed below, the variables for this analysis are the items and not the tasks.

## Requirements for data analysis

Classically, relationships between intraindividual variability on a set of variables are analysed by correlational methods and factor analysis. This approach can in principle identify the hierarchy of acquisitions predicted in the case of *universal décalages* (i.e., when the order of mastery is AB for all subjects), which should result in a simplex, and the local orders predicted by *individual décalages* (i.e., the order AB for certain subjects and the order BA for others, corresponding to different developmental pathways), which should result in group factors. A discussion of the relationships between these different types of *décalages* and the different types of factors identified by differentialists can be found in Longeot (1978). Although the present problem clearly calls for a multivariate analysis, classical factor analysis was not used in the present study since it is not the best way to handle the constraints imposed by the nature of the data and the issues at hand.

*Constraints arising from the qualitative nature of the data.* Most factor-analytic methods use correlations computed on variables assumed to be continuous and to have a normal distribution. These properties cannot be assumed to exist for variables operationalizing a theory which emphasizes structural changes and discontinuities over the course of development. Piagetian tasks are generally designed to induce one form of behavior if the operational structure is present, and another form of behavior if it has not yet emerged (intermediary responses in certain cases form a third form of behavior).

Another constraint in a study which aims at being both developmental and differential is to find a method which can both reveal potential hierarchical relations (associated with universal *décalages*) and potential equivalence relations (associated with individual *décalages*).

These constraints motivated the choice of a multivariate analysis which can handle qualitative data, and can reveal both equivalence and hierarchical relations. As shown below, correspondence analysis is equipped to handle these constraints.

*Constraints created by the necessity of establishing correspondence between item grouping and subject grouping.* In terms of relationships between variables, a multivariate analysis of interindividual differences in the form of intraindividual variability should result in the grouping of items having similar profiles (i.e., items being passed and failed by the same subjects).

In terms of relationships between subjects, this multivariate analysis should be capable of identifying clusters of subjects whose profiles are similar in terms of performance on items (i.e., subjects who succeed or fail on the same items).

Methods of data analysis can generally deal with one type of clustering or the other, but their simultaneous examination is often problematical. As its name suggests, the method used here preserves the *correspondence* between grouping of variables and grouping of subjects. Because of this feature, the developmental profile of those subjects contributing most to factors accounting for relations between variables can be identified easily.

*Constraints imposed by the longitudinal nature of the study.* Although correspondence analysis can identify individual *décalages* (i.e., interindividual differences in the form of intraindividual variability) at a given point in development, proof that these *décalages* correspond to developmental trajectories requires showing that they are stable in time. Correspondence analysis is applicable here too. It provides a means of plotting 'supplementary' individuals on the multivariate space defined by one analysis, who were not originally included in it. This feature means that the developmental profile of a subject tested later in time can be plotted on the initial analysis space. Comparing respective coordinate positions of individuals who were part of the analysis on the first occasion and then plotted as supplementary individuals on the second occasion, is one of the ways of testing the stability of a developmental profile.

The next section is devoted to a detailed description of the way in which correspondence analysis takes the constraints inherent in this study into account.

## CORRESPONDENCE ANALYSIS

Correspondence analysis has been popularized by Benzecri (1973, 1980), but see also Cibois (1983, 1984), Greenacre (1984), Greenacre & Hastie (1987), Escofier & Pagès (1988), Lebart, Morineau & Warwick (1984), Van der Heijden (1987). In this section a simple example is used to present the principles of correspondence analysis with a minimum of mathematical formulation. Readers familiar with the mathematics can, however, refer to the appendix which summarizes the equations referred to in the description.

Whereas classical factor analysis only accepts symmetric matrices, (correlation matrices), correspondence analysis can also handle non-symmetric matrices. More specifically, correspondence analysis exploits a matrix where the subjects are in rows and the tasks are in columns such that the number of times a subject succeeds on a task is expressed as a score on a row × column cross-tabulation. Since the task is only administered once, the matrix can only contain zeros or ones, which is illustrative of the qualitative nature of the data.

Take, for example, the fictitious matrix $T$ (see Table 9.1), where six individuals (rows) numbered from 1 to 6 were administered five tasks (columns) labelled A to E. The row × column table indicates a 1 if the subject succeeded and a 0 if s/he failed.

Correspondence analysis of this matrix serves a twofold purpose:

1. it decomposes the matrix into the sum of five particularly elementary matrices since they are obtained by simple multiplication of the marginal coefficients (five matrices termed 'one-dimensional' matrices since 5 is the smallest dimensionality of $T$);
2. it classifies these simplified matrices in descending order so as to be able to discard the latter. The magnitude of each matrix is indicated by the size of chi-square value.

Simply for the purposes of obtaining the above indicator, the initial matrix must be decomposed into the sum of several matrices. This is

Table 9.1. *Matrix T*

| | Tasks | | | | |
|---|---|---|---|---|---|
| Individuals | A | B | C | D | E |
| 1 | 0 | 1 | 0 | 1 | 0 |
| 2 | 1 | 1 | 0 | 1 | 0 |
| 3 | 1 | 1 | 1 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 1 |
| 5 | 0 | 0 | 1 | 0 | 1 |
| 6 | 0 | 1 | 0 | 1 | 1 |

Table 9.2. *Matrix $T_0$*

| | Tasks | | | | | |
|---|---|---|---|---|---|---|
| Individuals | A | B | C | D | E | |
| 1 | 0.2857 | 0.5714 | 0.2857 | 0.4286 | 0.4286 | 2 |
| 2 | 0.4286 | 0.8571 | 0.4286 | 0.6429 | 0.6429 | 3 |
| 3 | 0.4286 | 0.8571 | 0.4286 | 0.6429 | 0.6429 | 3 |
| 4 | 0.1429 | 0.2857 | 0.1429 | 0.2143 | 0.2143 | 1 |
| 5 | 0.2857 | 0.5714 | 0.2857 | 0.4286 | 0.4286 | 2 |
| 6 | 0.4286 | 0.8571 | 0.4286 | 0.6429 | 0.6429 | 3 |
| Total | 2 | 4 | 2 | 3 | 3 | 14 |

because calculating the difference between the observed value and the expected value under an independence assumption requires postulating that the original matrix is the sum of the matrix of the expected values and the matrix of the deviations from independence.

To return to our example, under the independence assumption the matrix $T_0$ is as in Table 9.2.

$T_0$ is already a simplified (one-dimensional) matrix since it is the product of the margins divided by the total. The matrix of deviations from independence is obtained by subtracting the independence matrix from the observed values. This yields the matrix $R_1 = T - T_0$ (see Table 9.3).

In this deviation from independence matrix $R_1$, the plus signs indicate success on tasks (scored 1) and the minus signs indicaté failure (scored 0). Thus the same qualitative information as in the initial matrix can be obtained purely by using signs.

Table 9.3. *Matrix $R_1$*

| | Tasks | | | | |
|---|---|---|---|---|---|
| Individuals | A | B | C | D | E |
| 1 | −0.2857 | 0.4286 | −0.2857 | 0.5714 | −0.4286 |
| 2 | 0.5714 | 0.1429 | −0.4286 | 0.3571 | −0.6429 |
| 3 | 0.5714 | 0.1429 | 0.5714 | −0.6429 | −0.6429 |
| 4 | −0.1429 | −0.2857 | −0.1429 | −0.2143 | 0.7857 |
| 5 | −0.2857 | −0.5714 | 0.7143 | −0.4286 | 0.5714 |
| 6 | −0.4286 | 0.1429 | −0.4286 | 0.3571 | 0.3571 |

Table 9.4. *Matrix $K_0$*

| | Tasks | | | | | |
|---|---|---|---|---|---|---|
| Individuals | A | B | C | D | E | Total |
| 1 | 0.2857 | 0.3214 | 0.2857 | 0.7619 | 0.4286 | 2.0833 |
| 2 | 0.7619 | 0.0238 | 0.4286 | 0.1984 | 0.6429 | 2.0556 |
| 3 | 0.7619 | 0.0238 | 0.7619 | 0.6429 | 0.6429 | 2.8333 |
| 4 | 0.1429 | 0.2857 | 0.1429 | 0.2143 | 2.8810 | 3.6667 |
| 5 | 0.2857 | 0.5714 | 1.7857 | 0.4286 | 0.7619 | 3.8333 |
| 6 | 0.4286 | 0.0238 | 0.4286 | 0.1984 | 0.1984 | 1.2778 |
| Total | 2.6667 | 1.2500 | 3.8333 | 2.4444 | 5.5556 | 15.7500 |

The chi-square value corresponding to these deviations is obtained as follows: the deviation from independence for each cell is squared and the result is weighted by the frequency corresponding to independence. This yields matrix $K_0$ (see Table 9.4).

In this matrix, the contribution to the chi-square values of the cells corresponding to failure are equal to the expected values, which is always the case when the observed values are null.

Factor decomposition in correspondence analysis is carried out by continuing the decomposing process begun with the independence matrix: the final goal is to find the one-dimensional matrix $T_1$ that exhibits the best fit with deviation matrix $R_1$ and at the same time accounts for the greatest contribution to the overall chi-square value.

To obtain the first matrix, the deviation matrix is subjected to an algorithm to search for the pair of eigenvectors defining this matrix (for a sample algorithm, see Cibois (1983)).

Table 9.5. *Matrix* $T_1$

| Individuals | Tasks A | B | C | D | E | $VI1$ |
|---|---|---|---|---|---|---|
| 1 | 0.1593 | 0.2747 | −0.1801 | 0.2425 | −0.4964 | 0.4391 |
| 2 | 0.2442 | 0.4211 | −0.2761 | 0.3716 | −0.7608 | 0.6730 |
| 3 | 0.0621 | 0.1071 | −0.0702 | 0.0945 | −0.1935 | 0.1711 |
| 4 | −0.1763 | −0.3041 | 0.1994 | −0.2683 | 0.5493 | −0.4859 |
| 5 | −0.2723 | −0.4696 | 0.3079 | −0.4144 | 0.8484 | −0.7504 |
| 6 | −0.0170 | −0.0293 | 0.0192 | −0.0259 | 0.0530 | −0.0469 |
| $VJ1$ | 0.3628 | 0.6258 | −0.4103 | 0.5522 | −1.1305 | |

Let $VJ1$ be the eigenvector corresponding to the columns of the matrix:

| $VJ1$ | A | B | C | D | E |
|---|---|---|---|---|---|
| | 0.3628 | 0.6258 | −0.4103 | 0.5522 | −1.1305 |

and $VI1$ be the eigenvector corresponding to the rows:

| $VI1$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| | 0.4391 | 0.6730 | 0.1711 | −0.4859 | −0.7504 | −0.0469 |

To obtain the first one-dimensional matrix $T_1$ approaching the deviations from independence matrix $R_1$, the elements of the eigenvectors are multiplied matricaly. For example, to obtain the value for the individual 1, task A, multiply 0.4391 by 0.3628, which yields 0.1593. In this fashion, matrix $T_1$ can be entirely reconstituted, which corresponds to this first factor (see Table 9.5).

Since the matrix $T_1$ covers part of the deviations from independence, it is possible to identify which part of the corresponding chi-square value it accounts for. This is done by calculating the contribution of each cell to the chi-square value of the matrix, and summing the rows, columns and the total.

For example, for reference cell (1, A) the part of the chi-square value is calculated by squaring the part of the deviation taken into account in the matrix and dividing the result by the original expected value. In other words, for this cell of matrix $K_1$ the values are: $0.1593 \times 0.1593/0.2857 = 0.0888$ (see Table 9.6).

The $K_1$ matrix is part of the original matrix $K_0$ which decomposes into $K_0 = K_1 + K_2 + K_3 + K_4$ (there are only four factors which contribute to the chi-square value because $T_0$, the independence matrix, makes no contribution).

Table 9.6. *Matrix* $K_1$

| Individuals | Tasks A | B | C | D | E | Total |
|---|---|---|---|---|---|---|
| 1 | 0.0888 | 0.1321 | 0.1136 | 0.1372 | 0.5749 | 1.0465 |
| 2 | 0.1391 | 0.2069 | 0.1779 | 0.2148 | 0.9004 | 1.6391 |
| 3 | 0.0090 | 0.0134 | 0.0115 | 0.0139 | 0.0582 | 0.1060 |
| 4 | 0.2176 | 0.3236 | 0.2782 | 0.3360 | 1.4082 | 2.5636 |
| 5 | 0.2594 | 0.3859 | 0.3317 | 0.4007 | 1.6793 | 3.0570 |
| 6 | 0.0007 | 0.0010 | 0.0009 | 0.0010 | 0.0044 | 0.0079 |
| Total | 0.7146 | 1.0628 | 0.9137 | 1.1036 | 4.6254 | 8.4201 |

Since this decomposition is additive, it can be seen that the contribution to the chi-square value of this first factor is from 8.4201 (the chi-square value for matrix $K_1$) to 15.75 (the chi-square value for matrix $K_0$), or 53%.

The factor decompound is thus still not sufficient since the decomposition of the first factor only accounts for 53% of the total chi-square value, and the procedure is reiterated. For this, matrix $T_1$ corresponding to the first factor is subtracted from the deviations from independence matrix $R_1$, and the remainder $R_2$ (not indicated here) is subjected to the algorithmic search for eigenvectors.

$$R_2 = R_1 - T_1$$

Let $VI2$ and $VJ2$ be the pair of eigenvectors corresponding to the second factor

| | | | | | | |
|---|---|---|---|---|---|---|
| $VJ2$ | A | B | C | D | E | |
| | −0.5126 | 0.1241 | −0.7497 | 0.6217 | 0.5165 | |
| $VI2$ | 1 | 2 | 3 | 4 | 5 | 6 |
| | 0.3758 | −0.0285 | −0.9465 | 0.2716 | −0.3197 | 0.6473 |

As was the case for the first factor, matrix $T_2$ which is the approximation of $R_2$ is reconstituted by multiplying the elements of the eigenvectors. Similarly, to determine what part of the chi-square value is accounted for by this factor, the chi-square value corresponding to this factor is obtained by squaring each cell and dividing by the initial expected value. The result shows that the sum of the contributions of all the cells of this second factor to the chi-square value is equal to 5.6280 or 36%: the total contribution of these two first factors to the chi-square value when summed is 89%.

Since nine-tenths of the chi-square value can thus be accounted for by the first two factors, the decomposition can stop here (the third factor would yield 99% and the fourth obviously 100%).

Because two factors are sufficient to cover the information contained in the deviation from independence matrix, these deviations can be displayed graphically where the $X$ axis gives the coordinates of the first eigenvector for individuals and for tasks, and the $Y$ axis indicates the second eigenvector: this yields the factorial graph shown in Figure 9.1.

This graph can be read in terms of angular conjunctions between rows and columns; for example, the vector from the origin to individual 1 who succeeded on tasks B and D is in angular conjunction with the vectors analogous to these tasks. In contrast, task A is in angular conjunction with individuals 2 and 3 who succeeded on this particular task.

This simultaneous representation of individuals and tasks makes it possible to establish correspondences between these two sets (hence the name correspondence analysis) and identify profiles of comparable individuals which can be accounted for by comparable successes. For example, individuals 1 and 2 both succeeded on tasks B and D, as can be seen by the reciprocal angular proximity of these four points.

The graph can also be interpreted globally. For example the first

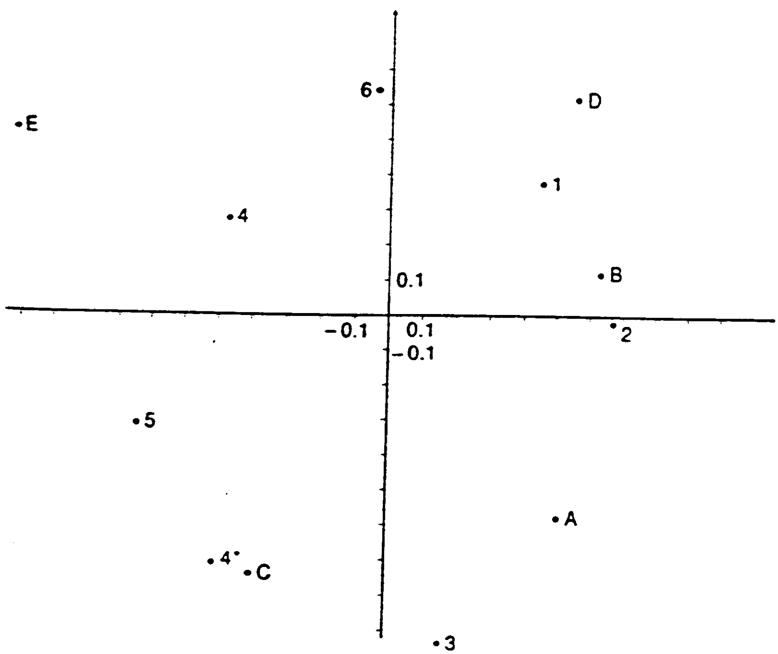


Figure 9.1. Factorial graph of the fictitious example.

(horizontal) axis contrasts success on tasks BDA to tasks EC, since the same individuals succeeded on each of these groups of tasks.

This finding can be verified easily by reclassifying the rows and columns according to the order of first factor: the opposition introduced by this factor is readily apparent.

| | B | D | A | C | E |
|---|---|---|---|---|---|
| 2 | 1 | 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 | 0 |
| 3 | 1 | 0 | 1 | 1 | 0 |
| 6 | 1 | 1 | 0 | 0 | 1 |
| 4 | 0 | 0 | 0 | 0 | 1 |
| 5 | 0 | 0 | 0 | 1 | 1 |

This reconstruction of the matrix clearly indicates the 'pure' individuals, i.e. all the subjects whose successes fall into one group of tasks and their failures into another. For instance individual 2 for BDA and individual 5 for CE are the closest to axis 1 and in addition are the individuals who contributed the most to the construction of this axis. If we return now to the chi-square table for the first factor ($K_1$) and insert the contributions of each of the rows to the total chi-square according to factor order we obtain:

| | |
|---|---|
| 2 | 1.6391 |
| 1 | 1.0465 |
| 3 | 0.1060 |
| 6 | 0.0079 |
| 4 | 2.5636 |
| 5 | 3.0570 |

This arrangement of the data clearly shows that the extremes on either side of the axis contribute the most to the chi-square value. In contrast individuals 3 and 6 who are the 'weakest' in terms of the contrast revealed by the first factor contribute the least.

These individuals who make no contribution to the first factor are however those who contribute the most to the second factor; a similar reconstitution of rows and columns can be obtained by following this factor order.

| | D | E | B | A | C |
|---|---|---|---|---|---|
| 6 | 1 | 1 | 1 | 0 | 0 |
| 1 | 1 | 0 | 1 | 0 | 0 |
| 4 | 0 | 1 | 0 | 0 | 0 |
| 2 | 1 | 0 | 1 | 1 | 0 |
| 5 | 0 | 1 | 0 | 0 | 1 |
| 3 | 0 | 0 | 1 | 1 | 1 |

This shows that the contrasts derive from two tasks and not three, which in addition explains why this factor is only in second position. It contrasts success on tasks DE ('pure' individual 6) with success on tasks AC ('pure' individual 3). A similar verification procedure shows that these individuals contribute the most to this factor.

The purpose of this example was to demonstrate that correspondence analysis can (1) treat categorized data in a purely descriptive manner, and (2) identify contrasts (factors) which reveal individual–task hierarchies. As a function of the interpretation assigned to a given factor, one hierarchy rather than another will be exploited. Note that the interpretation of contrasts is facilitated by the simultaneous representation of individuals and tasks. The profiles of individuals who contributed the most to each opposition can be identified.

Another feature of correspondence analysis is that it can be used to compare different kinds of individuals (or tasks): one possible application is longitudinal analysis. To return to the fictitious data, take the case of individual 4 who is retested on the same tasks at a later point in time. The results show greater mastery such that this subject now succeeds on tasks A and D as well as on task E as before. It would be useful to integrate this individual, denoted 4*, into the previous analysis without, however, having the data intervene directly. This is because the difference in time periods between analyses makes it unsound for us to associate these data with data from subjects tested at another time. On the other hand it would be of value to compare the results obtained by the same individuals at different points in time.

Correspondence analysis can be used to incorporate supplementary individuals into a completed analysis by determining the best plot fit to other individuals in the analysis with whom they share the greatest number of features. Individual 4* who succeeded on A, C and E is placed between individual 5 who succeeded on C and E, and individual 3 who succeeded on A and C.

The coordinates for 4* on the graph must be calculated factor by factor. For each factor, the coordinate for the individual is calculated by obtaining the algebraic sum of the factorial coordinates (divided by their marginal sum) and weighting this sum by the square root of the chi-square divided by the overall sum.

For individual 4* who succeeded on A, C and E this yields

| | Coordinate for the first factor | Frequency |
|---|---|---|
| A | 0.3628 | 2 |
| C | −0.4103 | 2 |
| E | −1.1305 | 3 |

The chi-square value for the first factor is 8.4201, and the overall sum

14. The algebraic sum of the coordinates divided by the sample size yields −0.4006 which is divided by the square root of (8.4201/14). This equals −0.5165 which will be the coordinate on the first factor for supplementary point 4*.

Obviously if a supplementary individual has exactly the same profile as an existing individual, the coordinates will overlay the same points. If we calculate a point for 1* who succeeded as 1 did, on B and C, it can be seen that the result is equal to 0.4391 which is the coordinate for 1 on the first factor. Numerically, this yields

$$\frac{((0.6258/4) + (0.5522/3))}{\sqrt{(8.4201/14)}} = 0.4391$$

Analogous calculations can be performed to plot 4* on the second factor; what changes is the factorial coordinates and the chi-square, the sums remains the same.

This method of calculation for supplementary individuals can also be used for tasks; in this case the coordinates of supplementary tasks are calculated by applying the same rules to the coordinates of individuals who performed on this task.

Other more sophisticated methods have recently been proposed to use correspondence analysis with longitudinal data. They emphasize the fact that correspondence analysis may be considered as a method of representation of residuals from expected values following the independence model.

It is possible to decompose residuals from model other than independence. For longitudinal data, Escofier & Pagès (1988) define the 'intra-analysis': their decomposition, using correspondence analysis, is from a model which suppresses the information that is not concerned with time (the inter-inertia) and shows the intra-inertia.

Van der Heijden (1987) decomposes residuals from log-linear models, from quasi-independence (when tables are incomplete because observations cannot possibly occur on given cells), from symmetry or quasi-symmetry. These methods may be used for contingency tables indexed by time, transition matrices, three and higher-wave univariate categorical panel data, multivariate categorical panel data, event-history data.

## RESULTS

Correspondence analysis was applied to the longitudinal data described in the first section. The 154 individuals tested on the first occasion appear in the rows and the 38 items they were administered appear in the columns. For each item, subjects were scored 1 if they succeeded or

0 if they failed (in fact, there are 76 columns, since success and failure are represented as two disjunctive modalities for each item).

## Analysis of the first occasion (cross-sectional approach)

Correspondence analysis of this matrix yields three interpretable factors (cf. Lautrey *et al.*, 1986).

*First factor.* In the plane formed by the first two factors (1 and 1 bis) the items' positions form a horseshoe which, in correspondence analysis, is one possible indication of a hierarchical relationship between variables, as a function of their rank on the first factor (in this configuration, factor 1 bis has a purely technical role and no psychological meaning: this is why this factor is not commented on and is labeled 1 bis). This hierarchical relationship is, however, very approximative and relative. Several groups of items have strong inner hierarchies, but these local hierarchies are only weakly interrelated. The first factor can be interpreted as a general factor of complexity (as regards items) and as a general factor of development (as regards subjects).

*Second factor.* The next factor contrasts 'logical' and 'infralogical' items. Piaget used the term 'logical' to refer to operations bearing on the relationships between *discrete objects* (the logical domain is hence discontinuous) and the term 'infralogical' to refer to operations bearing on relationships between *parts of objects* (the infralogical domain is continuous, for example space or time; the subject must thus isolate parts from the continuum before operating on them). Nevertheless, aside from this distinction, Piaget considered that logical and infralogical operations were isomorphic and arose from the same overall structure.

The simultaneous representation of items and subjects can be used to locate, on each of the two poles of the axis, the items and the individuals which contributed most to the part of the chi-square value that the axis accounts for. Reading Table 9.7 horizontally shows profiles of some of these individuals for these items, and vertically shows profiles of these items for these individuals.

The items are presented in the columns. Those which contribute most to the definition of the 'logical' pole of axis 2 appear on the left-hand side of the table and are denoted L. These items are tasks of varying difficulty and are about class intersection and quantification of probabilities. The items which contribute the most to the definition of the infralogical pole of axis 2 appear on the right-hand side of the table and have been labelled IL They cover tasks on the sectioning of volume,

Table 9.7. *Success patterns of the five subjects contributing the most to each pole of factor 2*

| Subjects | | | Logical items | | | | | | | Infralogical items | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sex | Age | | L1 | L2 | L3 | L4 | L5 | L6 | L7 | IL1 | IL2 | IL3 | IL4 | IL5 | IL6 |
| | | *N* | 100 | 76 | 68 | 44 | 40 | 18 | 18 | 64 | 45 | 32 | 39 | 24 | 19 |
| M | 7 | | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| M | 9 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| M | 9 | | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| F | 12 | | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| M | 10 | | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| F | 11 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| M | 12 | | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| F | 10 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| G | 12 | | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| F | 6 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |

folding of volumes and mental imagery. As in the fictitious example, the columns were reclassified within each group of items according to the order of the first factor. The numbers of the items (e.g., *L*1, *L*2, . . . , *Ln*) correspond to the order of their coordinates on this factor. The number of subjects *N* (out of 154) who succeeded on them appears below.

The subjects identified by sex (M or F) and age (6 to 12) are presented in the rows. The five subjects contributing the most to the 'logical' pole of axis 2 appear at the top and the five subjects who contributed the most to the 'infralogical' pole appear at the bottom. Within each of these groups the rows were reclassified as a function of the order of the coordinates on axis 2.

The shape of these patterns is entirely characteristic of what was termed 'individual *décalage*' or 'interindividual difference in the form of intraindividual variability' above. Some subjects apparently made progress in the logical domain while stagnating in the infralogical domain, whereas the reverse was observed for other subjects.

*Third factor.* The infralogical items which contributed most to the definition of the second factor are the tasks where the parts of objects that the individuals had to perform mental actions on were visible. The items contributing most to the third factor were infralogical items where the parts to be manipulated mentally could not be seen. Within this set of items, axis 3 contrasts items from the physical domain (e.g., conservation of volume) with items from the spatial domain (e.g., folds and holes). The table which can be derived from the items and subjects contributing the most to this factor exhibits the same shape as in Table 9.7.

## Longitudinal analysis

The correspondence analysis on the first evaluation is informative on the state of intraindividual *décalages* at a given point in development for each subject. However, to determine whether the *décalages* correspond to different trajectories in the course of cognitive development, it must be shown that these remain stable over time.

*Method of analysis of the relationships between the two evaluations.* The study of stability and changes in success profiles over time exploits the possibility of plotting supplementary individuals onto an analysis that they were not included in. The success profiles of subjects for the same set of items they were tested on three years previously (when they were 9–12 years of age) were plotted as supplementary individuals on the analysis of the first evaluation. The sample used in the first evaluation serves as an appropriate base of reference since it also treats subjects aged 9 to 12 who can be used for purposes of comparison. This procedure also has the additional advantage of situating each subject in terms of his/her own coordinate position three years earlier on an identical axis system.

*Stability and changes from evaluation* 1 *to evaluation* 2. (a) Stability and change in subject's absolute position. The metric on which correspondence analysis is based can be used to identify the distance between the two points characterizing a given subject on each evaluation, and to decompose this distance along the various axes. The coordinates of these two points are entirely comparable since the axis they are plotted on is the same. Rather than illustrate this feature in terms of subjects, we have opted to represent the coordinates for age groups on axis 1 for the two evaluations. These age groups have been treated here as supplementary individuals. The procedure for plotting the age groups is identical to the one used in the fictitious example for individual 4*, except that the profile for a fictitious individual representing an age group is obtained by averaging over the profiles of the subjects in this group.

Figure 9.2 illustrates axis 1 which is bounded at its extreme left by the coordinate of the easiest item ($I_1$) and on its extreme right by the most difficult item ($I_{39}$). The age group coordinates are indicated by arrows located above the axis for the first evaluation (A6 to A12), and below the axis for the second evaluation (A'9 to A'12). Groups A'9 to A'12 are thus made up of the same subjects as groups A6 and A9 three years

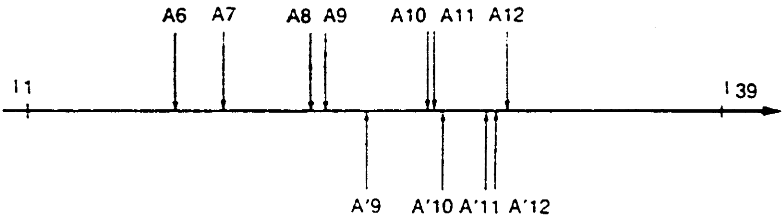


Figure 9.2. Simultaneous projection of items and individuals on the first axis. The 'individuals' are the age groups on the first (A6 to A12) and the second (A'9 to A'12) occasions.

earlier. Inspection of the plots of the age groups on the first occasion and their progress on the second occasion gives additional support to the assumption that the first factor is a general factor of development. The differences between age groups, however, are not all regular. One possible explanation is that development itself is irregular; another is that irregularities are due to problems in sampling subjects or variables. To clarify, changes in the coordinates of the groups over time can be examined. The short distance between groups A8 and A9 recurs three years later between groups A'11 and A'12, whereas the coordinates for groups A11 and A12 are spread normally. Similarly, the normal spread between A7 and A8 recurs between A'10 and A'11 whereas the distance between A10 and A11 is minimal. Thus in all likelihood the irregularities in the distances between age groups on axis 1 are due to problems of subject sampling. Note that the distances between groups are approximately the same from evaluation 1 to 2 along this axis.

(b) Stability and change in subjects' relative position. This aspect of stability and change can be assessed by correlations calculated for each axis between coordinates for individuals on the first evaluation (where they appear as main elements) and on the second evaluation (where they were plotted as supplementary elements). For the first four factors, the correlations are respectively 0.76 ($r$ significant at $p < 0.0001$), 0.35 ($p < 0.001$), 0.34 ($p < 0.002$) and −0.06 (NS).

These figures indicate that during the three-year time period, order of subject coordinates on the first factor remained fairly stable. In addition there is a weaker trend towards stability in intraindividual *décalages* which is accounted for by factors 2 and 3. This is shown by the level of significance but also by the difference in the value of the correlation on axis 2 or 3 and on axis 4 which could not be interpreted.

*Advantages and limitations of the 'supplementary individuals' technique.* These will be discussed by comparing the correlation values obtained by applying this method with correlations obtained through other methods.

The problems raised by the longitudinal comparison of two occasions can also be handled without resorting to supplementary elements. A second possibility is to perform a correspondence analysis on the matrix containing the profiles of the subjects in groups A6 to A8 and groups A'9 to A'11 (groups A9 and A'12 were dropped to avoid having two groups of nine year olds in the same analysis). In this case the subjects tested on the first and on the second occasion are incorporated in the same analysis. The third possibility is to perform two separate analyses, one on groups A6 to A9 and the second on groups A'9 to A'12. The comments that follow are restricted to the consequences of these choices on correlations between coordinates for the two evaluations on each of the first four factors. These are respectively 0.73 ($p < 0.0001$), 0.26 ($p < 0.05$), 0.21 ($p < 0.10$) and 0.08 (NS) if the second method is applied, and 0.70 ($p < 0.001$), −0.12 (NS), 0.21 ($p < 0.10$) and 0.23 ($p < 0.05$) for the third.

As shown by comparing the correlation values obtained using each of the three techniques, values are the highest when the 'supplementary individuals' technique is used to plot the success profiles for the second occasion on the factors identified in the analysis of the first. These correlations drop, mainly for factors 2 and 3, when the successes profiles for occasions 1 and 2 are analyzed together (second solution). They drop further, at least for factor 2, when the analyses are performed separately for occasions 1 and 2. In other words, the correlations between the coordinates on the factors for the two occasions are the highest when the second evaluation contributes the least to determining these factors.

This paradoxical finding suggests that the same meaning cannot be assigned to the factors identified on the first and second occasions. A detailed analysis of the items that contribute the most to the different factors show that this is indeed the case. Separate analyses of the 6–9 and 9–12 age groups on the first evaluation (the subjects used in the cross-sectional analysis) yield the same factors as those identified in the analysis of the entire sample (see Lautrey *et al.*, 1986). However, a separate analysis of the 9–12 age group on the second occasion identifies factors with a slightly different meaning, as indicated by the nature of the items located at each of the poles of the factors. The reason seems to be that the 9–12 age group on the second occasion is more advanced than the 9–12 age group on the first occasion (this can be seen in Figure 9.2, by comparing respectively the positions of groups A'9 and A9, A'10 and A10, A'11 and A11, on the first axis). This difference may be due to sampling fluctuations, or to the fact that subjects on the second occasion were taking the tests for the second time. The consequence is that those logical items in the class intersection task were no longer discriminant among these subjects. When this group is analyzed separately, discriminant logical items are no longer in sufficient number to give rise to a purely logical pole on the second factor. Logical items are thus mixed with infralogical items which were the nearest to them in the analysis of the first occasion. The change in the meaning of the factors between the two occasions is thus apparently due to technical rather than to theoretical reasons. In this case, the technique of plotting subject success profiles on the second evaluation as 'supplementary individuals' provides a means of constraining the factors used to analyze stability over time to conserve the same meaning. In the framework of this interpretation, it is no longer paradoxical that the correlations between the coordinates on the factors for the two occasions are the highest when the second occasion contributes the least to determining these factors. The technique of 'supplementary individuals' may not, however, be optimal in all circumstances: constraining the factors to preserve the meaning that they had on the first occasion only makes sense if the changes between the two occasions can be attributable to some undesirable artefacts.

## CONCLUSION

In the specific study which has served as an illustration here, correspondence analysis has been used to investigate individual differences in development as measured by a series of Piagetian tasks. More precisely, the aim was to analyze the structure and stability of interindividual differences in the form of intraindividual variability. The findings show that this variability is not entirely attributable to random fluctuations: the observed variability exhibits an interpretable structure and relative stability over time. This suggests that a multidimensional model of cognitive development may be better adapted than the unidimensional Piagetian model to account for observational data, including those obtained on Piagetian tasks. These data are congruent with the assumption that different trajectories are possible during cognitive development.

This example was selected because the constraints generated by both the theoretical issues and the nature of the data were particularly well suited to illustrating the potential of correspondence analysis for handling categorical data. The method has been shown here to be especially useful in cases where there is a need to perform a multivariate analysis of qualitative data. In this respect, correspondence analysis is comparable to multidimensional scaling. In addition it affords simultaneous representation of variable and individual space on the same axis system, which facilitates the analysis of correspondences between the structures observed in each of these two spaces, and provides complementary information on both. Lastly, correspondence analysis can situate supplementary elements in an analysis in which they were not originally included. This feature is doubtless the most valuable one for longitudinal studies. It can provide a useful solution to methodological problems that occur when the aim is to keep the position of axes constant across time occasions. The rationale for this solution is comparable to constraining the position of factors in the framework of confirmatory factor analysis. This kind of solution is naturally inapplicable when changes in the meaning of factors between two occasions are likely to have theoretical causes. In this case, it is better to compare factors in separate correspondence analyses or to use more sophisticated versions of correspondence analysis designed to handle longitudinal data, such as those suggested by Escofier & Pagès (1987) or by Van der Heijden (1987).

## APPENDIX

Number of subjects expressed as:

cell $n_{ij}$, line total $n_i$, column total $n_j$, grand total $n$

Frequencies: respectively

$$f_{ij} = n_{ij}/n, \quad f_i = n_i/n, \quad f_j = n_j/n$$

Conditional frequencies:

$$f_j^i = n_{ij}/n_i, \quad f_i^j = n_{ij}/n$$

Eigenvalues for a given factor are written:

$$\lambda \text{ and } \xi \text{ with } \xi = \sqrt{\lambda}$$

Eigenvectors calibrated, weighted: $y^i$ and $y^j$
Eigenvectors calibrated, unweighted: $y_i$ and $y_j$ where

$$y^j = y_j/f_j \quad \text{and} \quad y^i = y_i/f_i \tag{1}$$

Reconstitution equation for a given factor:

$$f_{ij} = y^i y^j f_i f_j / \sqrt{\lambda} = y_i y_j / \xi = (y_i/\sqrt{\xi})(y_j/\sqrt{\xi})$$

where $y_i/\sqrt{\xi}$ and $y_j/\sqrt{\xi}$ are termed semi-calibrated, unweighted eigenvectors. To reconstitute in terms of number of subjects and not in terms of proportions, the following equation is used:

$$n_{ij} = y_i y_j n / \xi = (y_i\sqrt{n}/\sqrt{\xi})(y_j\sqrt{n}/\sqrt{\xi})$$

The semi-calibrated eigenvectors written in terms of number of subjects are:

$$Y_i = y_i\sqrt{n}/\sqrt{\xi} \quad \text{and} \quad Y_j = y_j\sqrt{n}/\sqrt{\xi} \tag{2}$$

then $n_{ij} = Y_i Y_j$.

Transition equation (supplementary elements):

$$y^i = (1/\xi) \sum_j f_j^i / y^j$$

in a 0/1 matrix,

$$f_j^i = 1/n_i \quad \text{if} \quad n_{ij} = 1$$

and

$$f_j^i = 0 \quad \text{if} \quad n_{ij} = 0$$

then $\Sigma_j f_j^i y^j = (1/n) \Sigma_j y^j$

and

$$y^i = (1/n_i\xi) \sum_j y^j \qquad (3)$$

Equation (2) yields:

$$y_i = Y_i\sqrt{\xi}/\sqrt{n} \quad \text{and} \quad y_j = Y_j\sqrt{\xi}/\sqrt{n} \qquad (4)$$

Equation (1) yields:

$$y_i = y^i f_i \quad \text{and} \quad y_j = y^j f_j$$

where $f_i = n_i/n$ and $f_j = n_j/n$ which gives $y_i = y^i n_i/n$ and $y_j = y^j n_j/n$. $y_i$ is replaced by its value in (4):

$$Y_i\sqrt{\xi}/\sqrt{n} = y^i n_i/n$$

which yields

$$y^i = (Y_i\sqrt{\xi}/\sqrt{n})(n/n_i) = Y_i\sqrt{\xi}\sqrt{n}/n_i$$

and

$$y^j = Y_j\sqrt{\xi}\sqrt{n}/n_j$$

The equations make it possible to go from the results obtained in most programs ($y^i$ and $y^j$) to semi-calibrated values in terms of number of subjects ($Y_i$ and $Y_j$).

The values are entered into (3):

$$Y_i\sqrt{\xi}\sqrt{n}/n_i = (1/n_i\xi) \sum_j Y_j\sqrt{\xi}\sqrt{n}/n_j$$

which yields

$$Y_i = (1/\xi) \sum_j Y_j/n_j$$

similarly

$$Y_j = (1/\xi) \sum_i Y_i/n_i$$

Since the eigenvectors are expressed in terms of $\varphi^2$ and not in terms of $\chi^2$ where $\varphi^2 = \chi^2/n$,

If $\chi^2$ is the chi-square of a given eigenvector:

$$\xi = \sqrt{\chi^2}/\sqrt{n} \quad \text{and} \quad Y_i = (\sqrt{n}/\sqrt{\chi^2}) \sum_j Y_j/n_j$$

## REFERENCES

Benzecri, J. P. (1973). *L'analyse des données* (Vol. 2). Paris: Dunod.

Benzecri, J. P. (1980). *Pratique de l'analyse des données* (Vol. 1). Paris: Dunod.

Cibois, P. (1983). *L'analyse factorielle.* Paris: PUF.

Cibois, P. (1984). *L'analyse des données en sociologie.* Paris: PUF.

Escofier, B. & Pagès, J. (1988). *Analyses factorielles simples et multiples.* Paris: Dunod.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* New York: Academic Press.

Greenacre, M. J. & Hastie, P. (1987). The geometric interpretation of correspondence analysis. *Journal of the American Statistical Association, 82,* 437–447.

Lautrey, J., de Ribaupierre, A. & Riben, L. (1985). Intra-individual variability in the development of concrete operations: Relations between logical and infralogical operations. *Genetic, Social, and General Psychological Monographs, 111,* 167–192.

Lautrey, J., de Ribaupierre, A. & Rieben, L. (1986). Les différences dans la forme du développement cognitif évalué avec des épreuves piagétiennes: une application de l'analyse des correspondances. *Cahiers de Psychologie Cognitive, 6,* 575–613.

Lebart, L., Morineau, A. & Warwick, K. (1984). *Multivariate descriptive statistical analysis.* New York: Wiley.

Longeot, F. (1978). *Les stades opératoires de Piaget et les facteurs de l'intelligence.* Grenoble: Presses Universitaires de Grenoble.

de Ribaupierre, A., Rieben, L. & Lautrey, J. (1985). Horizontal decalages and individual differences in the development of concrete operations. In V. L. Shulman, L. C. Restaino-Baumann & L. Butler (eds.), *The future of piagetian theory: the neo-piagetians* (pp. 175–200). New York: Plenum Press.

Rieben, L., de Ribaupierre, A. & Lautrey, J. (1983). *Le développement opératoire de l'enfant entre 6 et 12 ans.* Paris: Editions du CNRS.

Van der Heijden, P. (1987). *Correspondence analysis of longitudinal categorical data.* Leiden: DWSO Press.